Memristor-based multilayer neural networks with online gradient descent training
Supplementary Material - Appendix

Daniel Soudry, Dotan Di Castro, Asaf Gal, Avinoam Kolodny, and Shahar Kvatinsky

February 8, 2015

A Generalization from memristors to memristive devices

In this paper, the memristor device is assumed to behave according to its classical model [1]. Though the first fabricated memristor device [2] has been modeled according to the classical model, this model is inaccurate and real devices can be modeled by the more general "memristive device" model. Furthermore, emerging memory technologies, *e.g.*, Resistive RAM and Spin-Torque Transfer MRAM, can be represented as memristive systems [3, 4]. A Memristive device [5] is a generalization of the original memristor [1]. For such devices the state variable can be a vector $\mathbf{s} \in \mathbb{R}^D$, and (assuming stationary dynamics)

$$\dot{\mathbf{s}} = \mathbf{f}(\mathbf{s}, u) \tag{1}$$
$$y = \psi(\mathbf{s}, u)\, u\,, \tag{2}$$

where u is the input (voltage/current) and y is the output (current/voltage). In [3], for example, s is a scalar as in the original memristor, but its dynamics is affected by some non-negative "window" function $\Theta(s)$

$$\dot{s} = f(u) \cdot \Theta(s)\ . \tag{3}$$

Usually $\Theta(s)$ is positive in some range and zero outside of that range (e.g., [3], Figs. 3-5). If s is in the range where $\Theta(s) = 0$, then $\dot{s} = 0$ for all times, and in this special case the memristive system is just a non-linear resistor. Therefore, one can safely assume that we start from a point $s(t=0)$ in which $\Theta(s(t=0)) > 0$. In that case, note that $\dot{s} \to 0$ near the edges of that range (where $\Theta(s(t)) \to 0$). Therefore, for all finite t, $s(t)$ cannot leave the range in which $\Theta(s(t)) > 0$. Thus one can (safely) define $z(s(t)) = \int_{s(0)}^{s(t)} (1/\Theta(x))\, dx$. By Leibniz rule, observe that

$$\dot{z} = (1/\Theta(s))\, \dot{s} = f(u)\ . \tag{4}$$

Additionally, since $z(s)$ is defined by an integral over a positive function, it is strictly monotone and therefore reversible to $s = h(z)$. Hence, the system can be represented as

$$\dot{z} = f(u) \tag{5}$$
$$y = \psi(h(z), u)\, u \tag{6}$$

Next, we consider the special case when $y = i$ (current), and $u = v$ (voltage). In this case we denote $\psi(h(z), v) = g(h(z), v)$ so it would be clear it represents conductance.

For a sufficiently small range of state space and inputs, $g(h(z), v)$ can be linearized around a fixed point (z^*, v^*), so (similarly to (3) in the paper)

$$g(h(z), v) = \bar{g} + \hat{g} z + \gamma v\,, \tag{7}$$

where

$$\begin{aligned}\hat{g} &\triangleq \left[\partial g\left(h\left(z\right),v^{*}\right)/\partial z\right]_{z=z^{*}}\\ \bar{g} &\triangleq g\left(h\left(z^{*}\right),v^{*}\right)-\hat{g}z^{*}-\gamma v^{*}\\ \gamma &\triangleq \left[\partial g\left(h\left(z^{*}\right),v\right)/\partial v\right]_{v=v^{*}}\,.\end{aligned}$$

Hence, a memristive system with a window function can be represented as

$$\dot{z} = f\left(v\right) \tag{8}$$

$$i = \left(\bar{g}+\hat{g}z+\gamma v\right)v\,. \tag{9}$$

Now this system is mathematically similar to the original system (1-3, in the paper).

Therefore, the only changes are the non-linearity in (8) and the γu correction in (9). In that case a similar method as in the original memristor case is used. Next, the required modification of the proposed design is described. Using the modified design, we again implement a synaptic grid circuit in a similar method as for memristors, with z replacing s as the synaptic weight.

Assume a sufficiently small input range in which f is reversible.

During the read cycle keep $u\left(t\right)=ax$ and replace the signal $\bar{u}\left(t\right)=-ax$ with $\bar{u}\left(t\right)=f^{-1}\left(-f\left(ax\right)\right)$. This modification is made so that the total change in the internal state variable is zero $\forall n,m$, since

$$\Delta z_{nm}=\int_{0}^{0.5T_{\mathrm{rd}}}f\left(ax_{m}\right)dt+\int_{0.5T_{\mathrm{rd}}}^{T_{\mathrm{rd}}}\left(-f\left(ax_{m}\right)\right)dt=0\,, \tag{10}$$

The output current of the synapse to the o_n line shortly after time zero is

$$I_{nm}=a(\bar{g}+\hat{g}z_{nm}+\gamma ax_{m})x_{m}. \tag{11}$$

Therefore, the total current in each output line o_n equals to the sum of the individual currents produced by the synapses driving that line, *i.e.*,

$$o_{n} = \sum_{m}I_{nm}=a\sum_{m}\left(\bar{g}+\hat{g}z_{nm}+\gamma ax_{m}\right)x_{m}\,. \tag{12}$$

The row output interface measures the output current o_n, and outputs

$$r_{n}=c\left(o_{n}-o_{\mathrm{ref}}\right) \tag{13}$$

where c is a constant converting the current units of o_n to a unit-less number r_n, and

$$o_{\mathrm{ref}}=\bar{g}a\sum_{m}x_{m}+\gamma a^{2}\sum_{m}x_{m}^{2}\,. \tag{14}$$

Note the a term $\gamma a^2\sum_m x_m^2$ was added to the reference signal o_{ref} to adjust for the extra γu term in (9). Defining

$$W_{nm}=ac\hat{g}z_{nm}\,, \tag{15}$$

we again obtain

$$\mathbf{r}=\mathbf{W}\mathbf{x}\,, \tag{16}$$

as desired.

During the write cycle replace the signals $u\left(t\right)=ax$ and $\bar{u}\left(t\right)=-ax$, respectively, with $u\left(t\right)=f^{-1}\left(ax\right)$ and $\bar{u}\left(t\right)=f^{-1}\left(-ax\right)$. This way the function $f\left(\cdot\right)$ in (8) is effectively 'canceled out'. As a result, we have

$$\dot{z}_{nm}=f\left(f^{-1}\left(a\mathrm{sign}\left(y_{n}\right)x_{m}\right)\right)=a\mathrm{sign}\left(y_{n}\right)x_{m}$$

so the total change in the internal state variable is exactly as we had in the original derivation (24, in the paper), $\forall n,m$:

$$\Delta z_{nm}=\int_{T_{\mathrm{rd}}}^{T_{\mathrm{rd}}+b|y_{n}|}\left(a\mathrm{sign}\left(y_{n}\right)x_{m}\right)dt=abx_{m}y_{n}\,.$$

If u is current and y is voltage in (5-6), a different design for the synapse should be used, as explained in the section A.1.

A.1 Current dependent memristive devices

As explained in section II.A in the paper, for a classical memristor the kinetics of the state variable can be treated either as voltage dependent or as current dependent. For a general memristive system (appendix A), however, this symmetry does not necessarily hold. It is possible that a memristor is only current dependent and not voltage dependent. In that case, changing the synaptic design is required as seen in Fig. 1a. In this case, (5-7) with $y = v$, and $u = i$, and $\psi(h(z), i) = 1/g(h(z), i)$ can be written as

$$\begin{aligned} \dot{z} &= f(i) \\ i &= g(h(z), i)\, v\,. \end{aligned}$$

Linearization with $\hat{g} = [\partial g(h(z), i^*)/\partial z]_{z=z^*}$ and $\bar{g} = g(h(z^*), i^*) - \hat{g}z^* - \gamma i^*$, $\gamma = [\partial g(h(z^*), i)/\partial i]_{v=v^*}$ yields

$$\begin{aligned} \dot{z} &= f(i)\,. \\ i &= (\bar{g} + \hat{g}z + \gamma i)\, v\,. \end{aligned} \tag{17}$$

Next, the operation of the system in Fig. 1 is described. The circuit contains four transistors, in addition to the memristor. During the operation of the circuit the M1 NMOS and the M2 PMOS function as a voltage controlled current sources. They both have low device parameter K (so they have a relatively low conductivity in comparison with the memristor) and therefore are always either in cutoff or in saturation. Also, both the M3 and M4 NMOS devices function as a transmission gate. They both have high device parameter K (so they have a relatively high conductivity in comparison with the memristor) and therefore are always either in cutoff or in the linear regime.

During the $T_{\rm rd}$-long read cycle, we have as depicted in Fig. 1b, $e_{\rm rd}(t) = V_{DD}$, $e_{\rm out}(t) = V_{DD}$, $u(t) = -\bar{u}(t) = -V_{DD}$ and

$$v_{\rm rd}(t) = \begin{cases} ax & ,\ \text{if } 0 \le t < 0.5T_{\rm rd} \\ -ax & ,\ \text{if } 0.5T_{\rm rd} \le t \le T_{\rm rd} \end{cases}\,. \tag{18}$$

In this case both M1 and M2 are at cut off, M3 and M4 are on (in the linear region), and with very high conductivity. Therefore, the voltage on the memristor is also $v_{\rm rd}(t)$. The current during the begining of the read procedure (at time 0^+) in each synapse is

$$I_{nm} \approx a(\bar{g} + \hat{g}z_{nm} + \gamma a\bar{g}x_m)x_m\,, \tag{19}$$

where we assumed that $\bar{g} \gg \hat{g}z_{nm} + \gamma a\bar{g}x_m$ (a small signal assumption). Additionally assuming $f(\cdot)$ is an odd function, using (17) and integrating over the read cycle we obtain

$$\begin{aligned} \Delta z_{nm} &= \int_0^{T_{\rm rd}} f(I_{nm}(t))\, dt \\ &\approx \int_0^{0.5T_{\rm rd}} f(a\bar{g}x_m)\, dt - \int_0^{0.5T_{\rm rd}} f(a\bar{g}x_m)\, dt \\ &= 0. \end{aligned}$$

Therefore, the read operation is approximately nondestructive (to zeroth order).

Therefore, the total current in each output line o_n equals to the sum of the individual currents produced by the synapses driving that line, *i.e.*,

$$o_n = \sum_m I_{nm} = a\sum_m (\bar{g} + \hat{g}z_{nm} + \gamma a\bar{g}x_m)\, x_m\,. \tag{20}$$

The row output interface measures the output current o_n, and outputs

$$r_n = c(o_n - o_{\rm ref}) \tag{21}$$

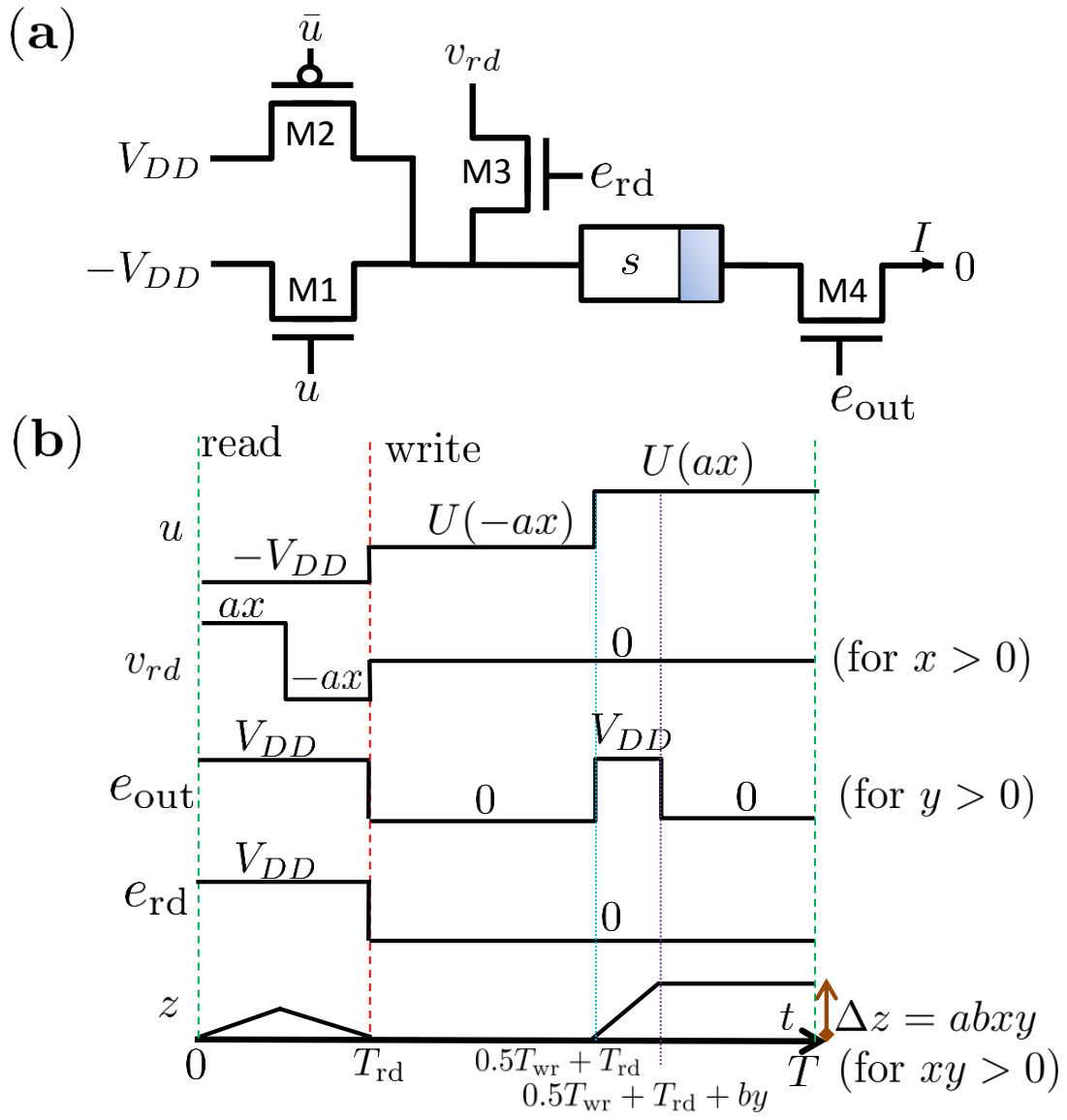

Figure 1: Current-based synapse design. **(a)** Schematic of the artificial synapse with input voltages $v_{\rm rd}, u$ and $\bar{u}$, control signals $e_{\rm out}$ and $e_{\rm read}$ and output current I. **(b)** Writing and reading protocol - incoming signals in a single synapse and the increments in the synaptic weight z.

where c is a constant converting the current units of o_n to a unit-less number r_n, and

$$o_{\text{ref}} = \bar{g}a \sum_m x_m + \gamma a^2 \bar{g} \sum_m x_m^2 \,. \tag{22}$$

Note the a term $\gamma a^2 \sum_m x_m^2$ was added to the reference signal o_{ref} to adjust for the extra γu term in (9). Defining

$$W_{nm} = ac\hat{g}z_{nm} \,, \tag{23}$$

we again obtain

$$\mathbf{r} = \mathbf{W}\mathbf{x} \,, \tag{24}$$

as desired.

During the write cycle $e_{\text{rd}}(t) = 0$ and

$$u(t) = \begin{cases} U(-ax) & , \text{ if } 0 \le t - T_{\text{rd}} \le 0.5T_{\text{wr}} \\ U(ax) & , \text{ if } 0.5T_{\text{wr}} < t - T_{\text{rd}} < T_{\text{wr}} \end{cases} , \tag{25}$$

$$\bar{u}(t) = \begin{cases} \bar{U}(-ax) & , \text{ if } 0 \le t - T_{\text{rd}} \le 0.5T_{\text{wr}} \\ \bar{U}(ax) & , \text{ if } 0.5T_{\text{wr}} < t - T_{\text{rd}} < T_{\text{wr}} \,, \end{cases} \tag{26}$$

where

$$U(\alpha) \triangleq \begin{cases} \sqrt{f^{-1}(\alpha)/K} + V_T - V_{DD} & , \text{ if } f^{-1}(\alpha) < 0 \\ -V_{DD} & , \text{ if } f^{-1}(\alpha) \ge 0 \end{cases} \tag{27}$$

$$\bar{U}(\alpha) \triangleq \begin{cases} -\sqrt{f^{-1}(\alpha)/K} - V_T + V_{DD} & , \text{ if } f^{-1}(\alpha) \ge 0 \\ V_{DD} & , \text{ if } f^{-1}(\alpha) < 0 \end{cases} \tag{28}$$

where K is the device parameter of transistors M1 and M2. Additionally, $e_{\text{out}}(t) = V_{DD}$ if

$$\min(by, 0) \le t - T_{\text{rd}} - 0.5T_{\text{wr}} \le \max(by, 0) \tag{29}$$

and zero otherwise. When $e_{\text{out}}(t) = V_{DD}$, it enables current flow through the memristor. In that time either M1 is saturated and M2 is cutoff, or M1 is cutoff and M2 is saturated.

For example, during $0.5T_{\text{wr}} < t - T_{\text{rd}} < T_{\text{wr}}$, if $f^{-1}(ax) < 0$, M2 is cutoff and M1 is saturated. The current on the memristor, arriving from M1, is

$$\begin{aligned} I &= -K(V_{GS} - V_T)^2 \\ &= -K\left(\sqrt{f^{-1}(ax)/K} + V_T - V_{DD} + V_{DD} - V_T\right)^2 \\ &= -\left|f^{-1}(ax)\right| . \end{aligned}$$

Also, during the same time, if $f^{-1}(ax) \ge 0$, M1 is cutoff and M2 is saturated. The current on the memristor, arriving from M1, is

$$\begin{aligned} I &= -K(V_{GS} - V_T)^2 \\ &= K\left(-\sqrt{f^{-1}(\alpha)/K} - V_T + V_{DD} - V_{DD} + V_T\right)^2 \\ &= \left|f^{-1}(ax)\right| . \end{aligned}$$

When $0 \le t - T_{\text{rd}} \le 0.5T_{\text{wr}}$ we just need to flip the sign of the ax argument. Taking into account all these cases, the current on the memristor is during the write cycle

$$I = \begin{cases} f^{-1}(-ax) \,. & , \text{ if } 0 \le t - T_{\text{rd}} \le 0.5T_{\text{wr}} \\ f^{-1}(ax) & , \text{ if } 0.5T_{\text{wr}} < t - T_{\text{rd}} < T_{\text{wr}} \,. \end{cases}$$

Therefore, when $e_{\text{out}}(t) = V_{DD}$, combining this with (17) yields

$$\dot{z}(t) = f(I) = \begin{cases} -ax\,. & , \text{ if } 0 \leq t - T_{\text{rd}} \leq 0.5T_{\text{wr}} \\ ax & , \text{ if } 0.5T_{\text{wr}} < t - T_{\text{rd}} < T_{\text{wr}}\,. \end{cases}$$

and zero otherwise. Integrating $\dot{z}$ over both the write cycle (note the current can only flow at times given in (29))

$$\begin{aligned} \Delta z &= \begin{cases} \int_{T_{\text{rd}}}^{T_{\text{rd}}+by} axdt & , \text{if } y \geq 0 \\ -\int_{T_{\text{rd}}-by}^{T_{\text{rd}}} axdt & , \text{if } y < 0 \end{cases} \\ &= \text{sign}(y) \int_{T_{\text{rd}}}^{T_{\text{rd}}+b|y|} axdt \\ &= abxy \end{aligned}$$

as desired.

B Direct voltage multiplication

The circuit proposed in this paper, implements a multiplication using (pulse duration)$\times$(signal strength). This novel method is used since direct multiplication of voltage/current signals is difficult to accurately execute with a small number of simple components [6]. Such an approximate method for a direct multiplication of voltage signals is shown in Fig. 2. This alternative design should be used if the memristor conductance is much higher then the transistor conductance, in contrast to our assumption (13, in the paper).

Consider a classical **current**-dependent memristor with dynamics as in Eq. 17, where for simplicity we assume that $f(i) = i$ and $\gamma = 0$. We denote by $R(z) = \bar{r} + \hat{r}z$ the state-dependent resistance of the memristor. Assume that for transistors M1 and M2 the threshold voltage is zero[1] $V_T = 0$, that $by \gg ax$ and that the transistors gain K is set sufficiently low, so that $ax \gg R(z)\,I$. Note that since the threshold voltage is zero, this means that both transistors are either in the linear region or in cut-off. Applying the voltages as shown in Fig. 2 (for $y, x > 0$), the current that flows through the memristor during the write cycle is

$$\begin{aligned} I &= K\left(V_{GS} - V_T\right) V_{DS} - 0.5V_{DS}^2 \\ &= K\left(by - R(z)\,I - V_T\right)\left(ax - R(z)\,I\right) \\ &- 0.5\left(\left(ax - R(z)\,I\right)\right)^2 \\ &\approx Kabxy\,, \end{aligned} \tag{30}$$

and similarly for all the other x, y quadrants, $I \approx Kabxy$. Denoting $\eta = KabT_{\text{wr}}$, and integrating over the write cycle yields

$$\Delta z = \eta xy\,, \tag{31}$$

as desired. Note that in this design K is low, while the conductivity of the memristor is high, which is the opposite case to the assumption in (13, in the paper). Additionally, this direct voltage multiplication method has similarity with the multiplication method suggested in [7] for Hebbian learning in CMOS synapses. In contrast to the method suggested here, the result of the multiplication in [7] must be positive, which makes it unusable for practical algorithms.

[1] Note there are CMOS transistors with zero threshold voltage (and even negative voltage for NMOS). For example in depletion-mode MOSFET a channel exists even with zero voltage.

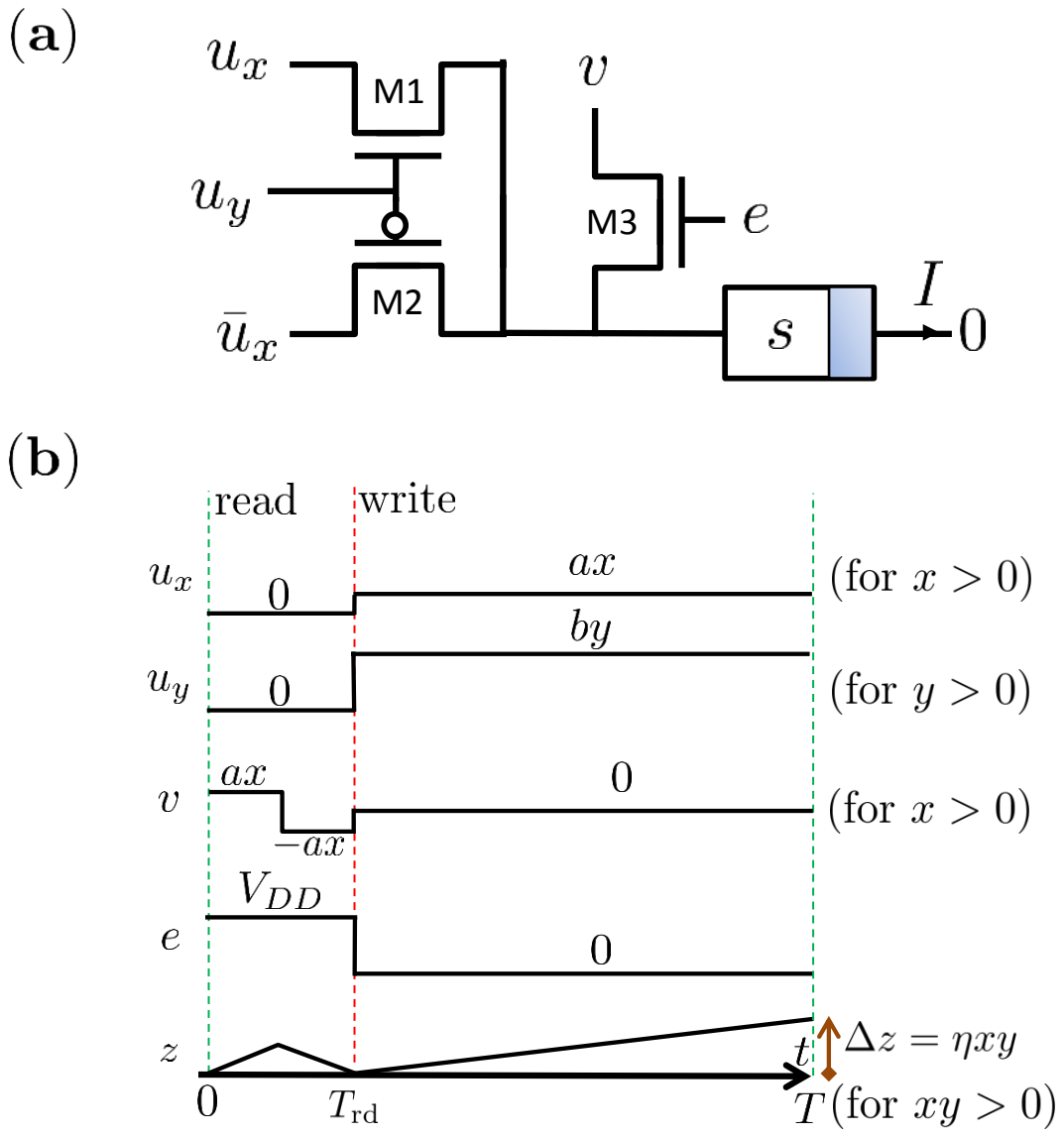

Figure 2: **Direct voltage multiplication synapse (a)** Schematic of the artificial synapse with input voltages v, u_y, u_x and $\bar{u}_x = -u_x$, control signal e, and output current I. **(b)** Writing and reading protocol - incoming signals in a single synapse and the increments in the synaptic weight s.

C Compact synapses

It is possible to reduce the number of transistors in each synapse from two to one, if one is willing to double the write time in the original device (thus slowing the operation of the circuit). The schematic of such a synapse is shown in Fig. 3a.

For simplicity, assume a classical memristor as in (1-2, in the paper).

As depicted in Fig. 3b, the read cycle is performed by applying, for a T_{rd} duration,

$$u\left(t\right) = \begin{cases} ax & , \text{ if } 0 \leq t < 0.5T_{\mathrm{rd}} \\ -ax & , \text{ if } 0.5T_{\mathrm{rd}} \leq t \leq T_{rd} \end{cases}, \tag{32}$$

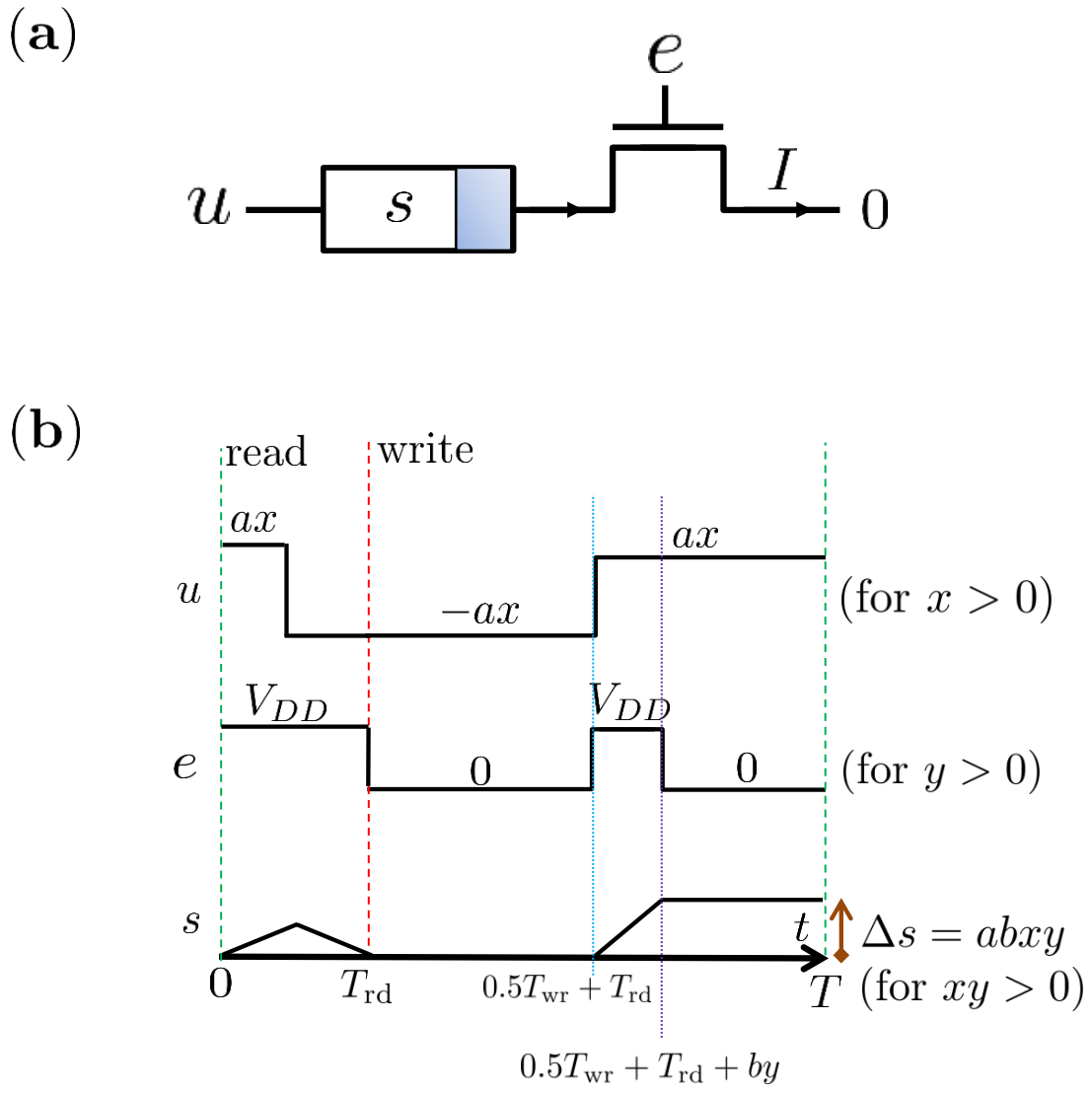

Figure 3: Compact synapse design. **(a)** Schematic of the artificial synapse with input voltages u control signal e and output current I. **(b)** Writing and reading protocol - incoming signals in a single synapse and the increments in the synaptic weight s

and $e(t) = V_{DD}$, so $\dot{s}(t) = u(t)$ and $\Delta s = 0$ over the read cycle. Sampling the current at the beginning of the read cycle gives

$$I = a(\bar{g} + \hat{g}s)x, \tag{33}$$

as required.

In the write cycle

$$u(t) = \begin{cases} -ax & , \text{ if } T_{\text{rd}} \le t \le T_{\text{rd}} + 0.5T_{\text{wr}} \\ ax & , \text{ if } T_{\text{rd}} + 0.5T_{\text{wr}} < t < T \end{cases}, \tag{34}$$

and $e(t) = V_{DD}$ if

$$\min(by, 0) \le t - T_{\text{rd}} - 0.5T_{\text{wr}} \le \max(by, 0) \tag{35}$$

and zero otherwise. Therefore, $\dot{s}(t) = ax$ if $e(t) = V_{DD}$, and zero otherwise.

Integrating over both the write half cycles, we obtain again

$$\Delta s = abxy\,. \tag{36}$$

as required.

D Spice simulation of circuit

D.1 Linear ion drift model

The proposed synapse array was tested using CMOS 0.18 μm process and linear ion drift memristor model [2, 3]. The test was set similarly to Fig. 6 in the paper, on a small 2×2 synaptic grid circuit (without the second read cycle), simulated for time $10T$ with simple inputs

$$(x_1, x_2) = (1, -2) \cdot 10\text{sign}\,(\text{t} - 5\text{T}) \tag{37}$$

$$(y_1, y_2) = (0.5, -0.25)\,. \tag{38}$$

and parameters as follows

- Memristors: $R_{\text{ON}} = 100\,\Omega, R_{\text{OFF}} = 100k\Omega$, $D = 10\,nm$ and $\mu_v = 10^{-14} m^2/(\text{s} \cdot V)$.
- Timing: $T = 0.1\text{sec}, T_{\text{wr}} = 0.6T$.
- Scaling: $a = 1mV, b = 0.6T, c^{-1} = 0.01A$.
- Power supply: $V_{DD} = 1.8V$.

As can be seen in Fig. 4 the circuit exhibited similar results as in Fig. 6 in the paper.

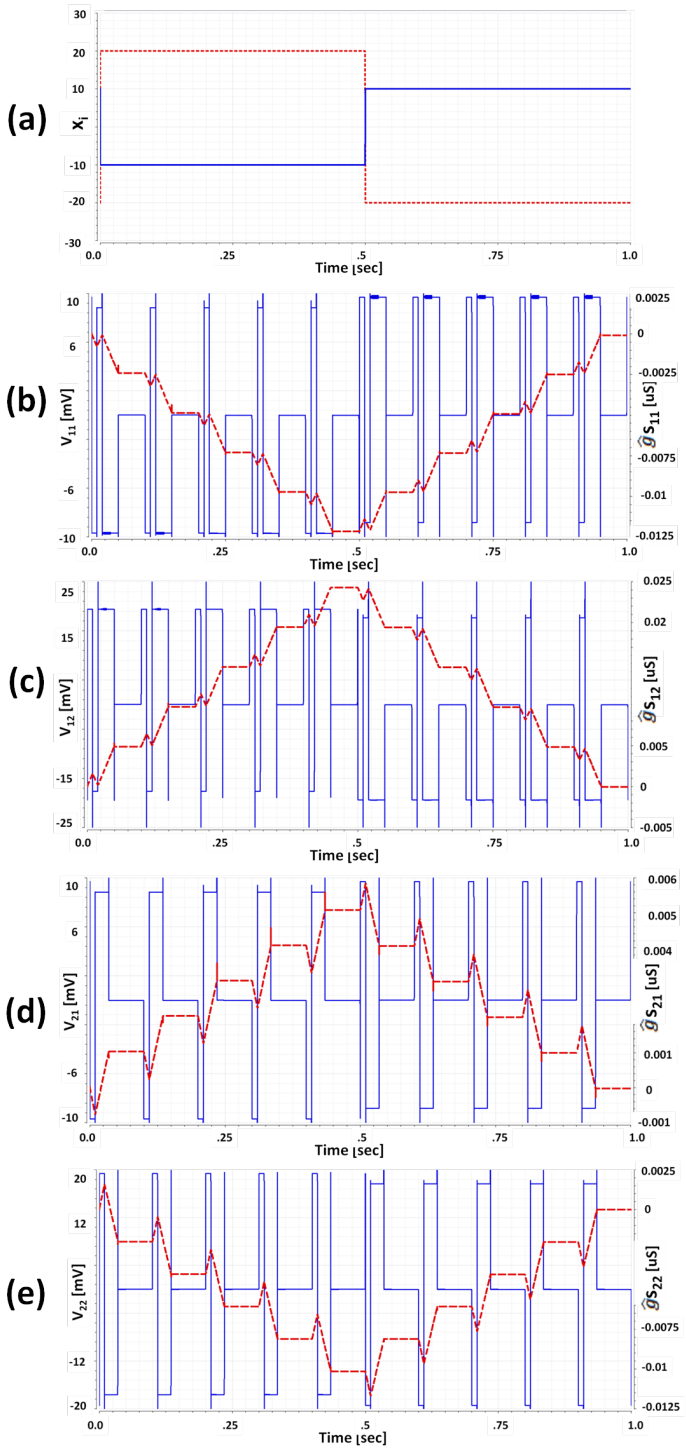

Figure 4: SPICE simulation of the proposed synapse for 0.18 μm CMOS process with linear ion drift memristors [2, 3]. **(a)** u_1 and u_2 wave forms. The voltage V upon the memristor and the conductance g of the memristor are shown, respectively, by the blue solid and red dashed lines in: **(b)** V_{11} and $\hat{g}s_{11}$, **(c)** V_{12} and $\hat{g}s_{12}$, **(d)** V_{21} and $\hat{g}s_{21}$, and **(e)** V_{22} and $\hat{g}s_{22}$. The input of the circuit and the parameters are explained in the text (appendix D.1).

D.2 Threshold adaptive memristor (TEAM) model

The proposed synapse design was tested using the TEAM memristor model [3], which fits well to practical memristive devices. The following parameters were used: $R_{\mathrm{ON}} = 100\,\Omega, R_{\mathrm{OFF}} = 200k\Omega$, Biolek window ($p = 2$), $k_{\mathrm{OFF}} = -k_{\mathrm{ON}} = 10$, $\alpha_{\mathrm{ON}} = \alpha_{\mathrm{OFF}} = 54$, $D = 3$ nm and the transistor was again modeled using CMOS 0.18 μm process.

The test was set again on a small 2×2 synaptic grid circuit, simulated for time $10T$ with ($T = 60\mu s$) with simple inputs and an analog control circuit. As before, the conductance can be adjusted (Fig. 5) using the read and write scheme described in the paper.

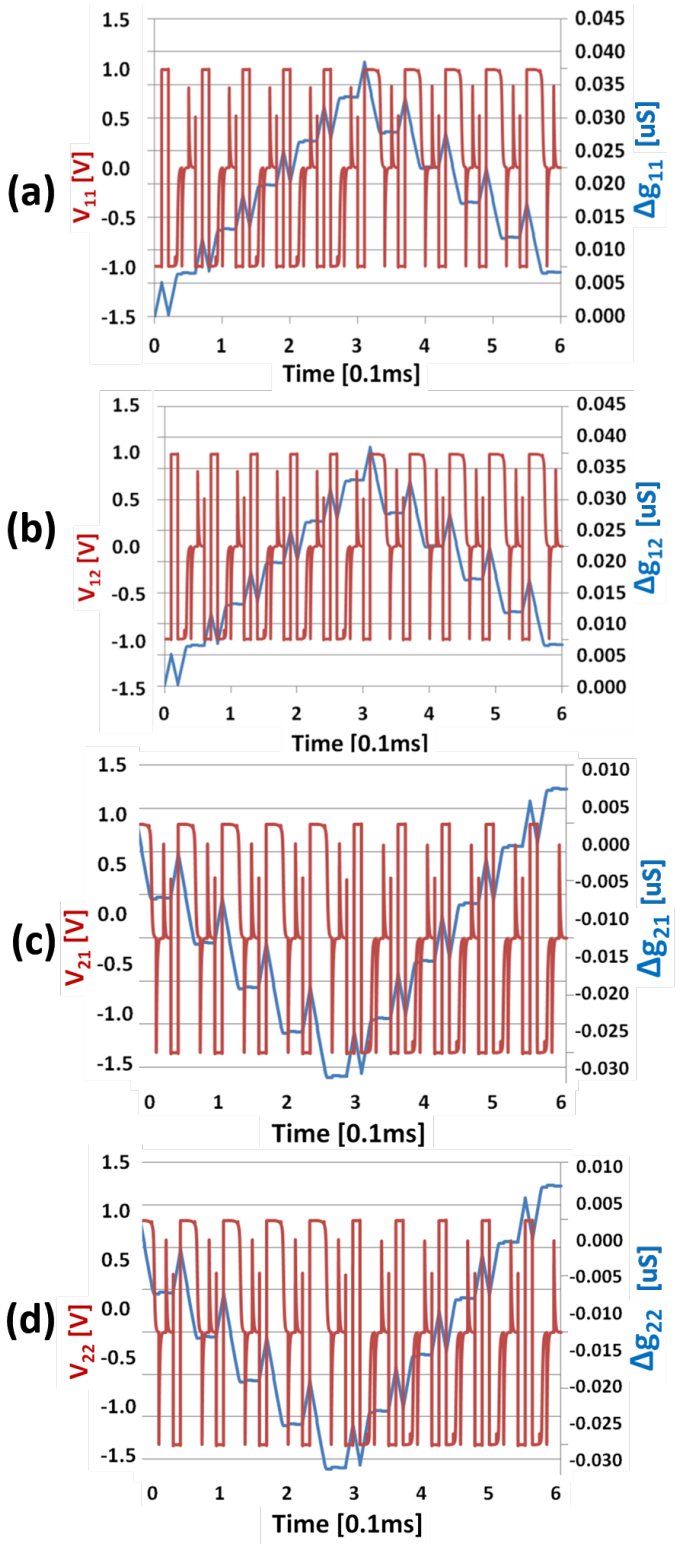

Figure 5: SPICE simulation of the proposed synapse for 0.18 μm CMOS process with TEAM model memristors [3]. The voltage V upon the memristor and the conductance change in the memristor are shown, respectively, by the blue solid and red lines in: **(a)** V_{11} and Δg_{11}, **(c)** V_{12} and Δg_{12}, **(d)** V_{21} and Δg_{21}, and **(e)** V_{22} and Δg_{22}. More details appear in the text (appendix D.2).

References

[1] L. Chua, "Memristor-the missing circuit element," *Circuit Theory, IEEE Transactions on*, vol. 18, no. 5, pp. 507–519, Sep. 1971.

[2] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80–83, 2008.

[3] S. Kvatinsky, E. G. Friedman, A. Kolodny, and U. C. Weiser, "TEAM: ThrEshold Adaptive Memristor Model," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 1, pp. 211–221, Jan. 2013.

[4] L. Chua, "Resistance switching memories are memristors," *Applied Physics A*, vol. 102, no. 4, pp. 765–783, 2011.

[5] L. Chua and S. M. Kang, "Memristive devices and systems," *Proceedings of the IEEE*, vol. 64, no. 2, 1976.

[6] G. Han and E. Sanchez-Sinencio, "CMOS transconductance multipliers: A tutorial," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, no. 12, pp. 1550–1563, 1998.

[7] H. Card, C. R. Schneider, and W. R. Moore, "Hebbian plasticity in mos synapses," *IEEE Transactions on Audio and Electroacoustics*, vol. 138, no. 1, pp. 13–16, Feb. 1991.